Data Quality in a SaaS Environment

Diby Malakar
VP – Engineering Operations, Cloud9 Analytics

Agenda

- What is SaaS?
  - 5 Upcoming Trends
  - Key Characteristics
  - Key Drivers behind adoption
- Impact of SaaS to BI and Analytics
  - Is it for real?
  - Future Adoption Trends
  - Top Challenges
- DQ in a SaaS Environment
- Q & A

What is SaaS?

- SaaS stands for Software As a Service
- SaaS is a
  - Software application delivery model where a software vendor develops a web-native software application
  - Hosts and operates (either independently or through a third-party) the application for use by its customers over the Internet

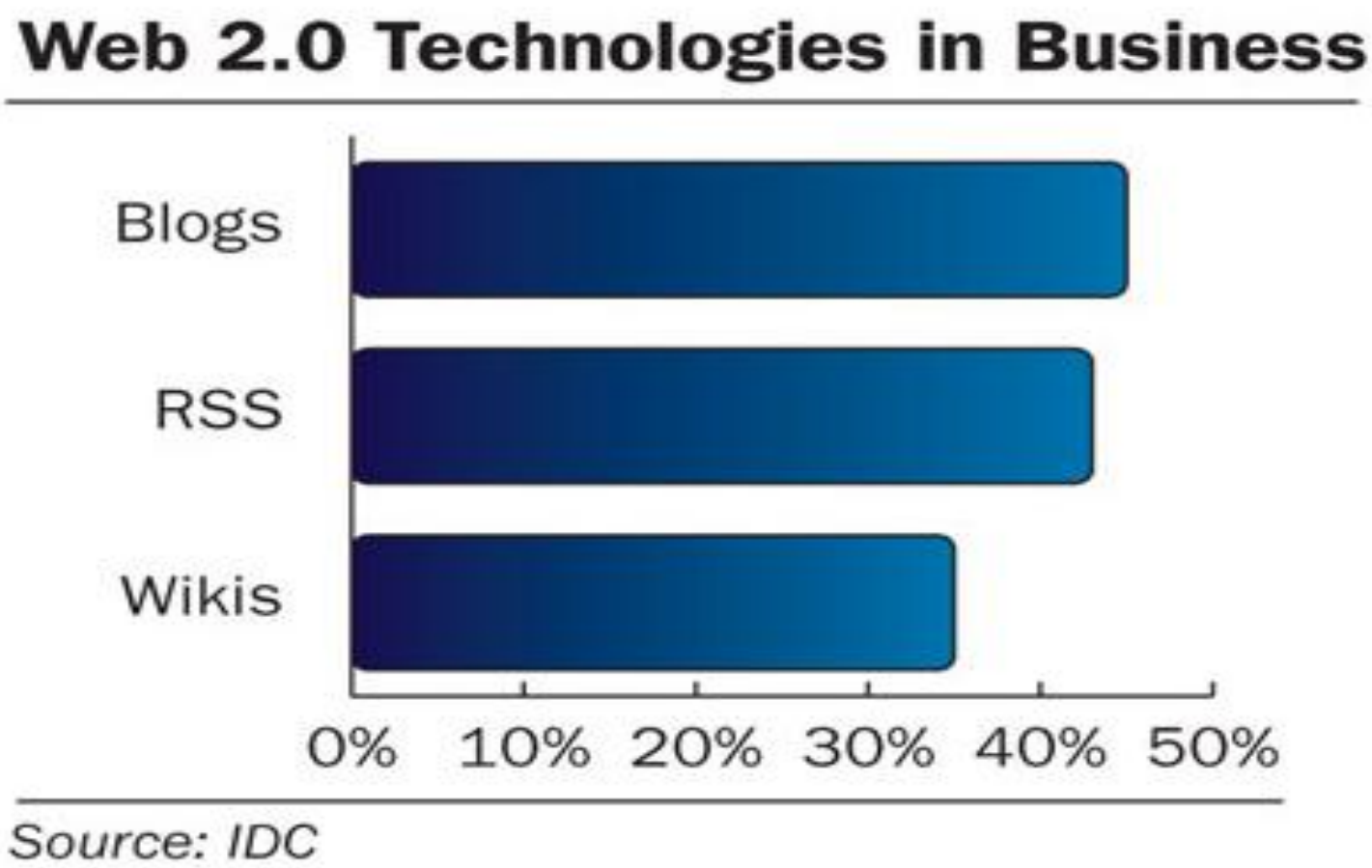
Source: http://en.wikipedia.org/wiki/Software_as_a_Service

5 Upcoming Trends

- *Software as a Service (SaaS)*
- Consumerization
- Web 2.0
- [Free and] Open Source Software
- Global Class Architectures

Source: Gartner's 5 Discontinuities for IT

Usage of Web 2.0 Technologies



Web 2.0 Technologies

Web 2.0 Tools Used by Large Enterprises		
Software Type	Software Providers	Corporate Adopters
Blog an individual's online journal	TypePad® Blogger (Google service) WordPress (free software) Movable Type®	General Motors Hitachi Infosys Intel Macromedia Novell Sun Microsystems
Wiki users create and edit content	MediaWiki TWiki Kwiki PmWiki Socialtext	Dresdner Kleinwort Microsoft Nokia SAP
RSS really simple syndication	RSS feed reader RSS specifications HexaMail	Amazon Cisco The Wall Street Journal
Tagging use of keywords to track content	ConnectBeam	HoneywellIBMSony-BMG
Social Networking use of the internet to build and maintain relationships	Humsubka Jhoom LinkedIn® Minglebox	Cisco Dresdner Kleinwort Microsoft Nike Salary.com
Mashup Web site or Web application that combines content from more than one source	Greasemonkey	Amazon Dun & Bradstreet E*TradeGoogle IBM JDS Uniphase Siemens Société Générale
Prediction Markets speculative markets created to improve forecasting		Google HP Microsoft Yahoo!
Source: KPMG		

What it means for IT?

"Information technology is no longer under the control of the IT organization"

- Tom Austin, Gartner Group analyst

The relationship between users and central IT - a "civil war"

Dealing with IT discontinuities

- Provide interactive tools like RSS Feeds, blogs, wiki's
- Help customers innovate (themselves) and foster independence
- Allow experimentation with new software and communities
- Segment users into categories based on experience

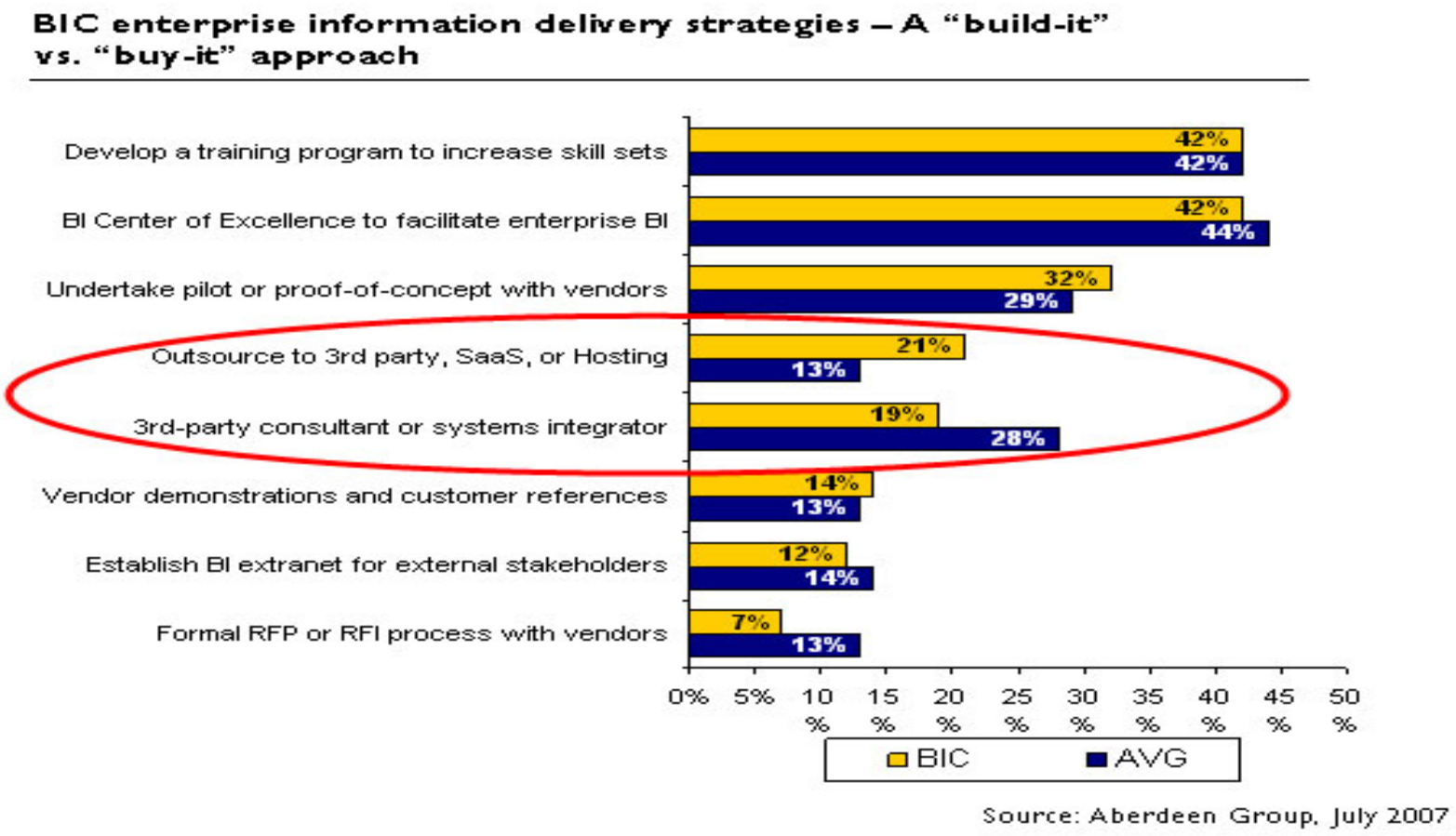
Key Characteristics of SaaS

- Centralized feature updates
- Single-instance, multi-tenant architecture
- Managed centrally and accessed over the internet
- Generally priced on a per-user basis
- Mostly subscription-based, no upfront license costs
- Short implementation cycle

Key Drivers behind adoption

- Most people use standardized apps
- New apps built using business logic atop common app platform from parameters and macros
- Reliability and popularity of web applications
- Security is sufficiently well-trusted (*)
- Improved network bandwidth
- Low Cost of ownership

SaaS and BI


BIC enterprise information delivery strategies – A "build-it" vs. "buy-it" approach
Develop a training program to increase skill sets
42%
42%
BI Center of Excellence to facilitate enterprise BI
42%
44%
Undertake pilot or proof-of-concept with vendors
32%
29%
Outsource to 3rd party, SaaS, or Hosting
21%
13%
3rd-party consultant or systems integrator
19%
28%
Vendor demonstrations and customer references
14%
13%
Establish BI extranet for external stakeholders
12%
14%
Formal RFP or RFI process with vendors
7%
13%
0% 5% 10% 15% 20% 25% 30% 35% 40% 45% 50%
BIC
AVG
Source: Aberdeen Group, July 2007

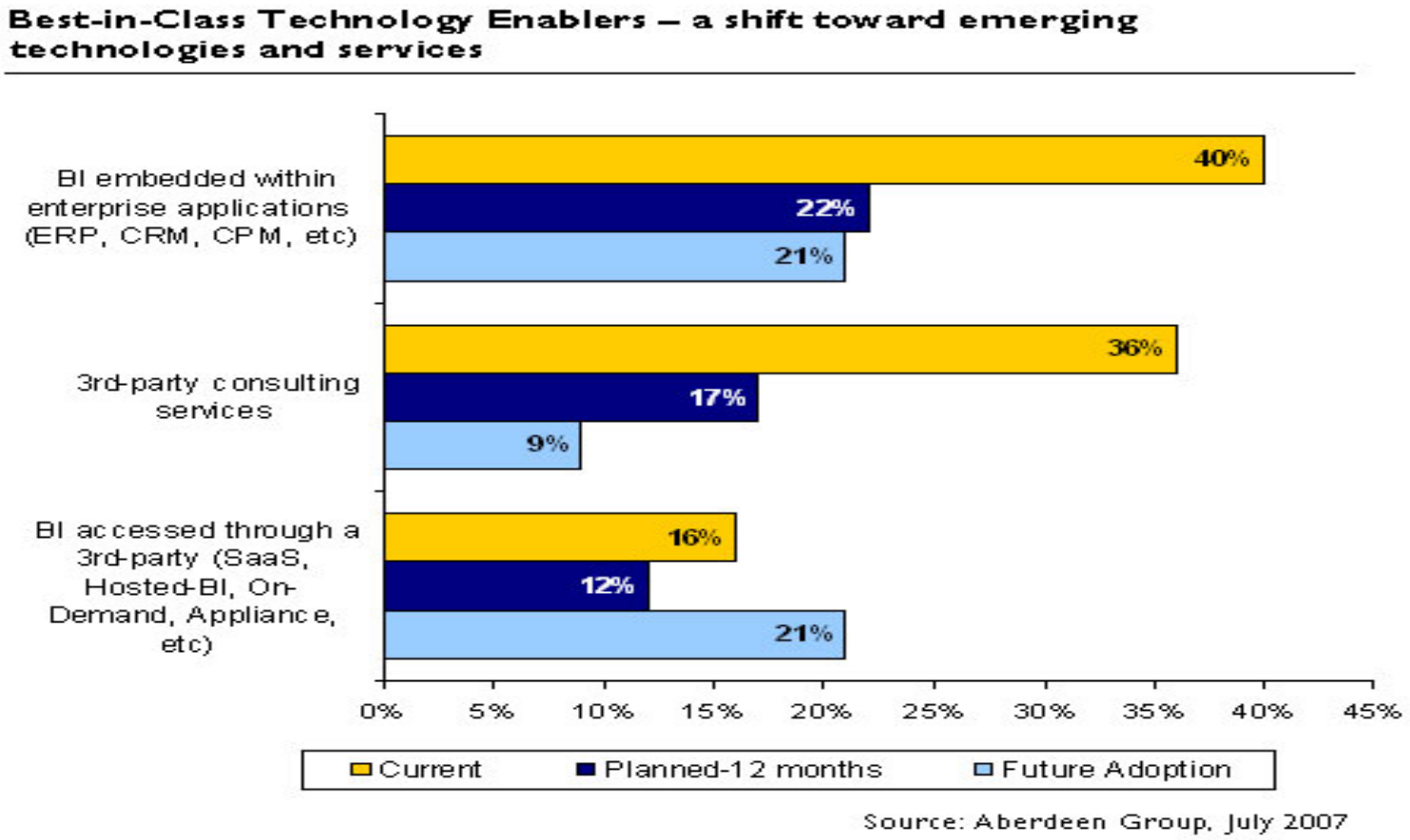
Gartner on SaaS and BI

- Consider SaaS-based BI if you have
  - A straightforward set of requirements around scheduled reporting
  - Monitoring a particular piece of data
  - Handling repetitive analysis of X measures

Primary Usage: To drive down Cost of Ownership and Cost of Skills…

"Business Value at a lower price point"

Future adoption of SaaS based BI



Top Challenges in SaaS BI

- Data Quality
  - Most vendors do not consider this as part of their service
- Metadata Quality
  - Can be an issue across multiple applications
- Customization can have limitations
  - Iterative Ad hoc reporting and high interactivity
- BI Support is different from OLTP support
  - Explaining numbers that don’t look right…customer may have different data patterns

SaaS implementation in DQ

- Current primary focus is on "Pro-Active" Data Quality Assessment (DQA)

DQA is defined as …

Inspection, **Measurement**, and **Analysis** of data to help business users understand the defects in the data and the impact of those defects upon the business

DQ Assessment Philosophy

- Involves
  - Identification of inaccurate, inconsistent or duplicate records
  - Focus on DQ than s/w, h/w and admin issues, no IT
  - Scheduled data assessment (no operators needed)
  - Profiling and Reporting - DQA Dashboard to list top data issues
  - Pattern recognition

Philosophy : *"Upfront data analysis"* makes it easier to conduct critical downstream cleansing

Q & A

Speaker Contact Info

Diby Malakar
VP – Engineering Operations, Cloud9 Analytics

diby.malakar@cloud9analytics.com
dibymalakar@hotmail.com

Cell: 408-375-2570